

Preliminary Measurement of Communication Rates on the Cray T3D Interprocessor Network

**Paul Springer
John Peterson**

(Jet Propulsion Laboratory, California Institute of Technology)

**Robert Numrich
(Gray Research, Inc.)**

Task Description

Objectives

- **Test the communication network which connects nodes on the Cray T3D**
- **Determine how communication rates scale with message size**
- **Test the effect of path length on transfer rates**
- **Show the effect of contention**

Accomplishments

- **Set of benchmark tests developed and run**
- **Transfer rates and latency times plotted**
- **Charts showing how communication rates scale with message size and path length**
- **Results of tests with and without network contention**

Hardware Overview

Nodes

- Two processing elements (PE) per node
- Each PE has its own processor (DEC Alpha), local memory, and support circuitry

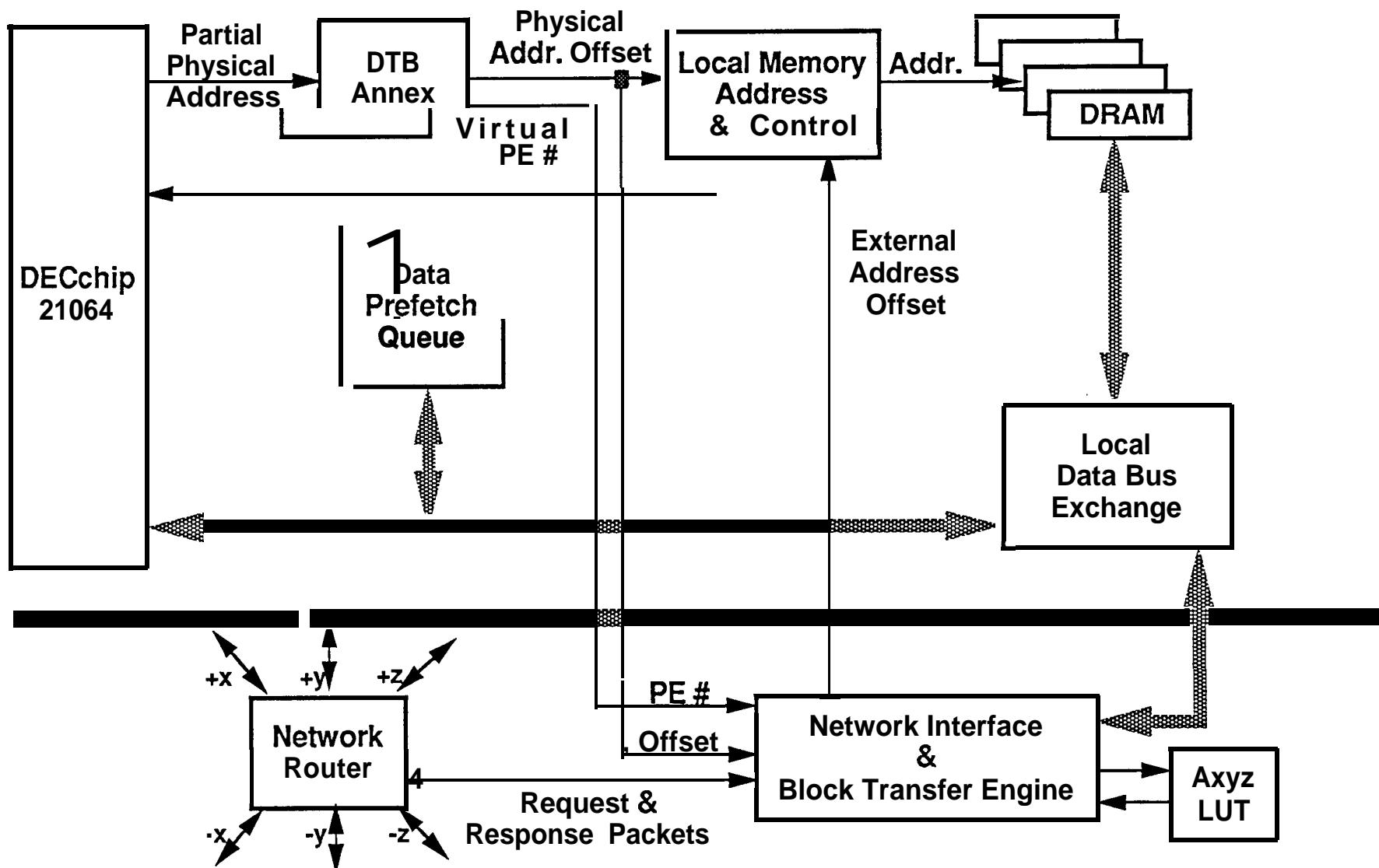
Memory

- Physically distributed across PEs
- Globally addressable via a single address space
- Memory access faster for local memory than memory in remote PEs

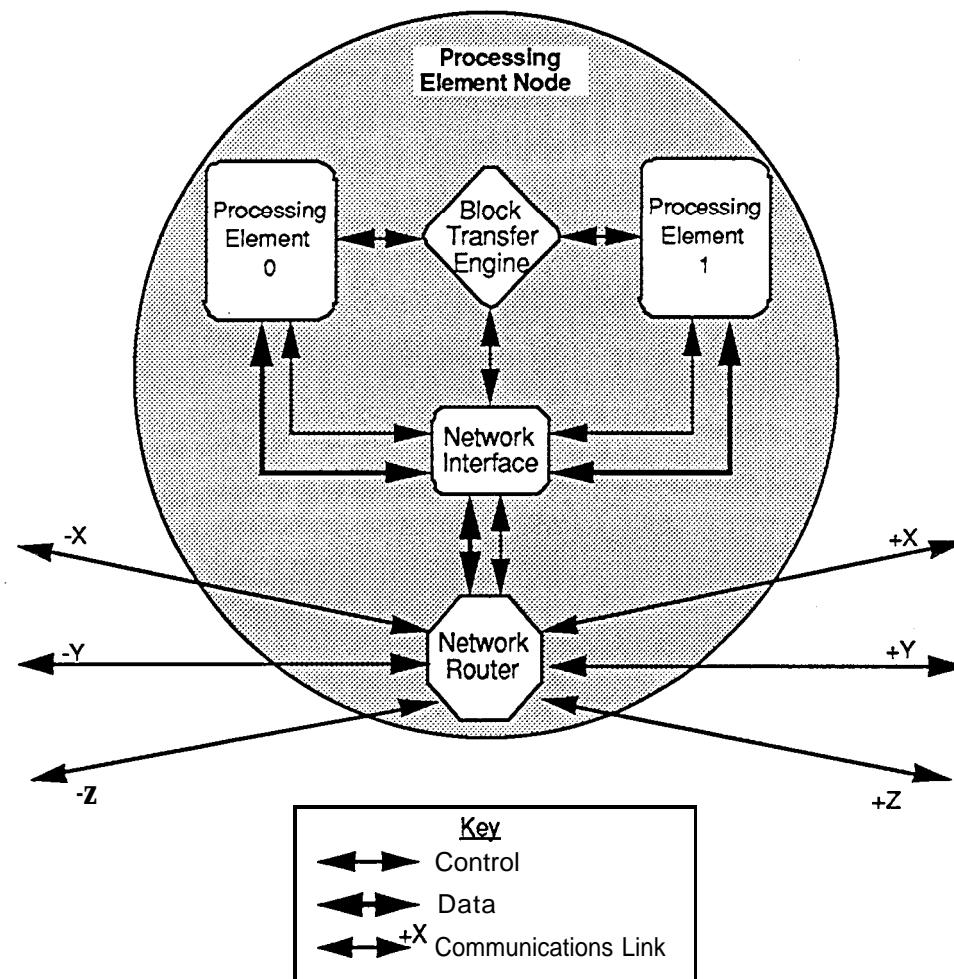
Communication Network

- 3-D Torus bidirectional interconnect network
- Operates asynchronously and independently of PEs
- Peak interprocessor communications rates of 300 Mbytes/ second in each dimension

PE Internal Organization

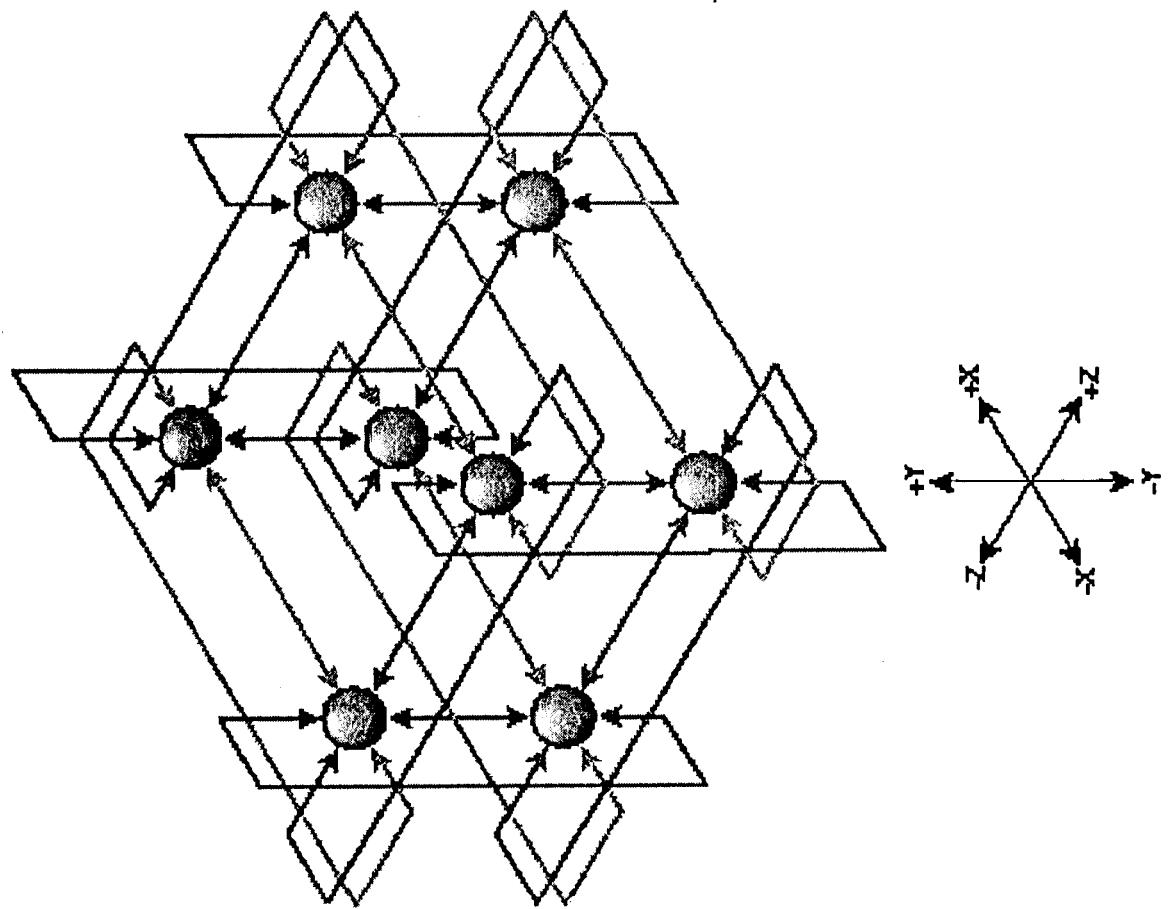


Interconnect Network Components



Preliminary Measurement of Communication Rates on the Cray T3D Interprocessor Network

3-D Torus



Interconnect Network Routing

- **Data moves across network in packets**
- **Each packet contains a routing tag**
- **Routing tag contains all necessary routing information, including number of hops in each dimension**
- **Routing tag generated by network interface, which uses a static routing table loaded at boot time**
- **Dimension order routing X dimension traversed first, then Y, then Z**

Interconnect Network: Addressing

- **Memory accesses make use of “DTB annex”**
- **DTB annex consists of a table located in the PE support circuitry**
- **Each DTB annex entry consists of a PE number and a function code**
- **Five bits of address specify which DTB entry to use**
- **The network interface translates virtual PE numbers to logical PE numbers**

Benchmark Description

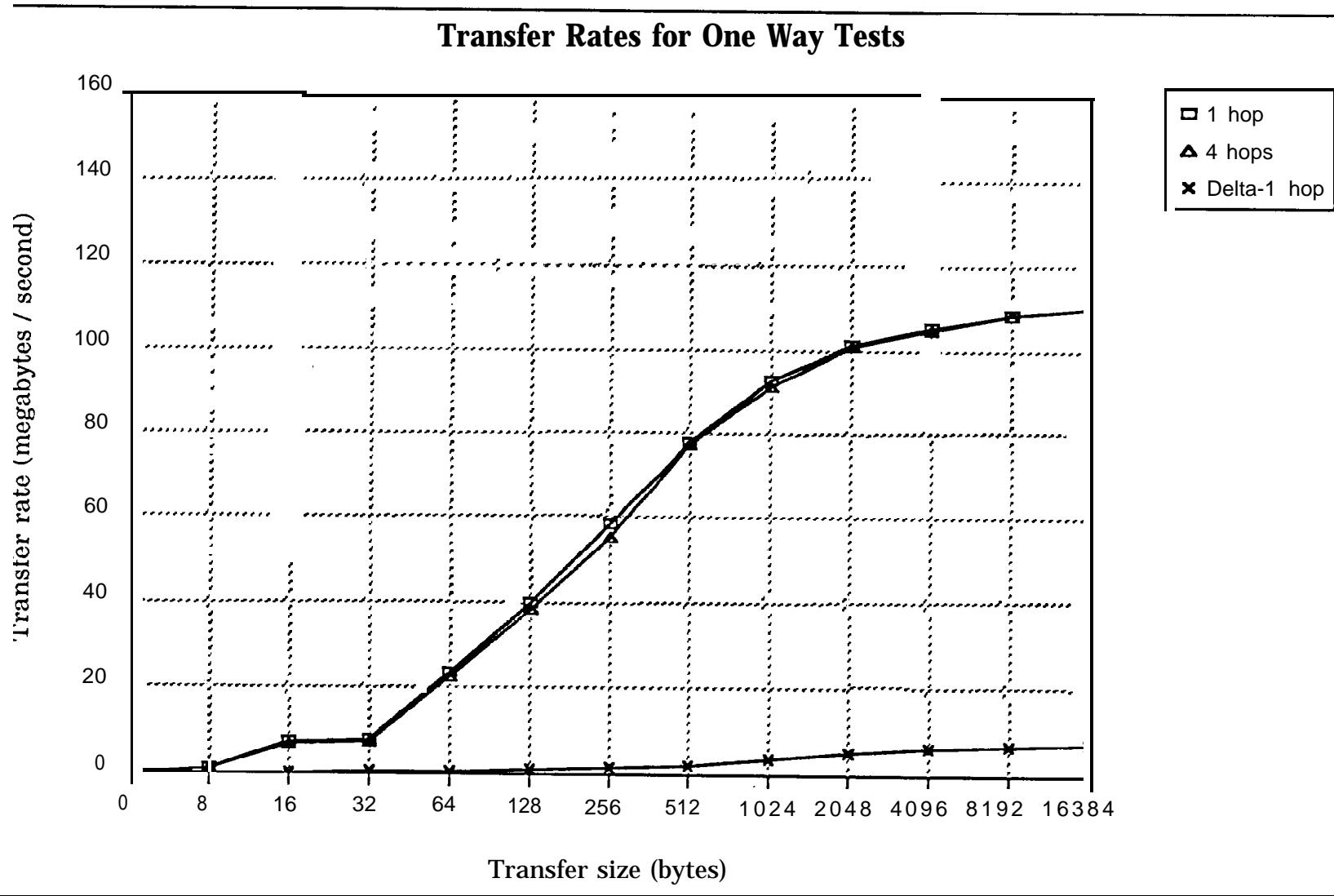
One Way Tests

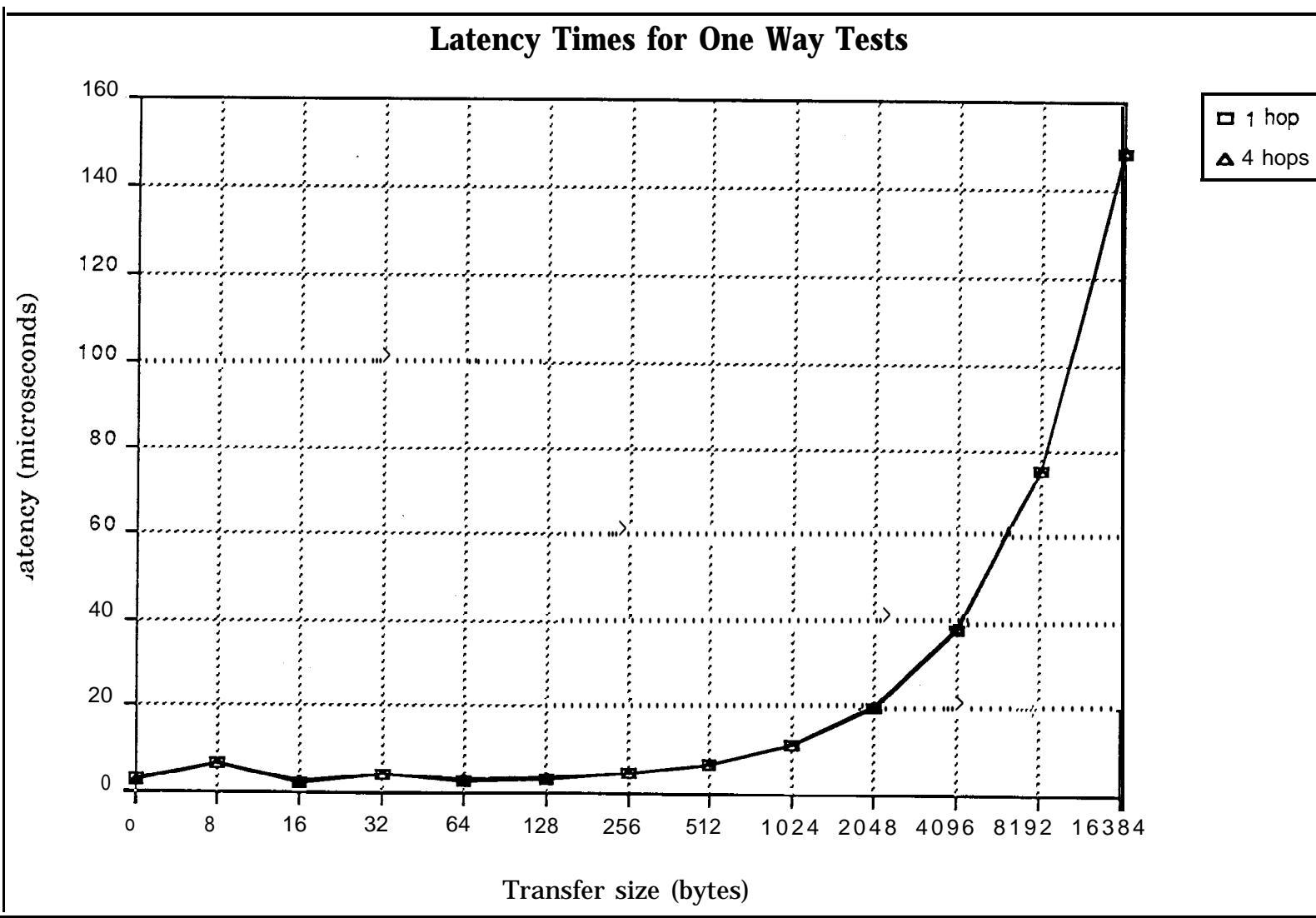
- **One PE writes to the memory of another PE on a separate node**
- **Remote PE sends acknowledgment packets back to first PE**
- **No contention: one write done at a time**
- **Timing figures shown include waiting for acknowledgment packets**

Results

- **Transfer rate of 110 megabytes/ second for 16K transfers,
compared to 7.2 megabytes/ second on the Delta***
- **Latency times as low as 2.3 microseconds for 16 byte transfer**
- **Longer path made very little difference in timing**

* Delta figures from a private communication from Dr. Roy Williams, Caltech, October 18, 1993





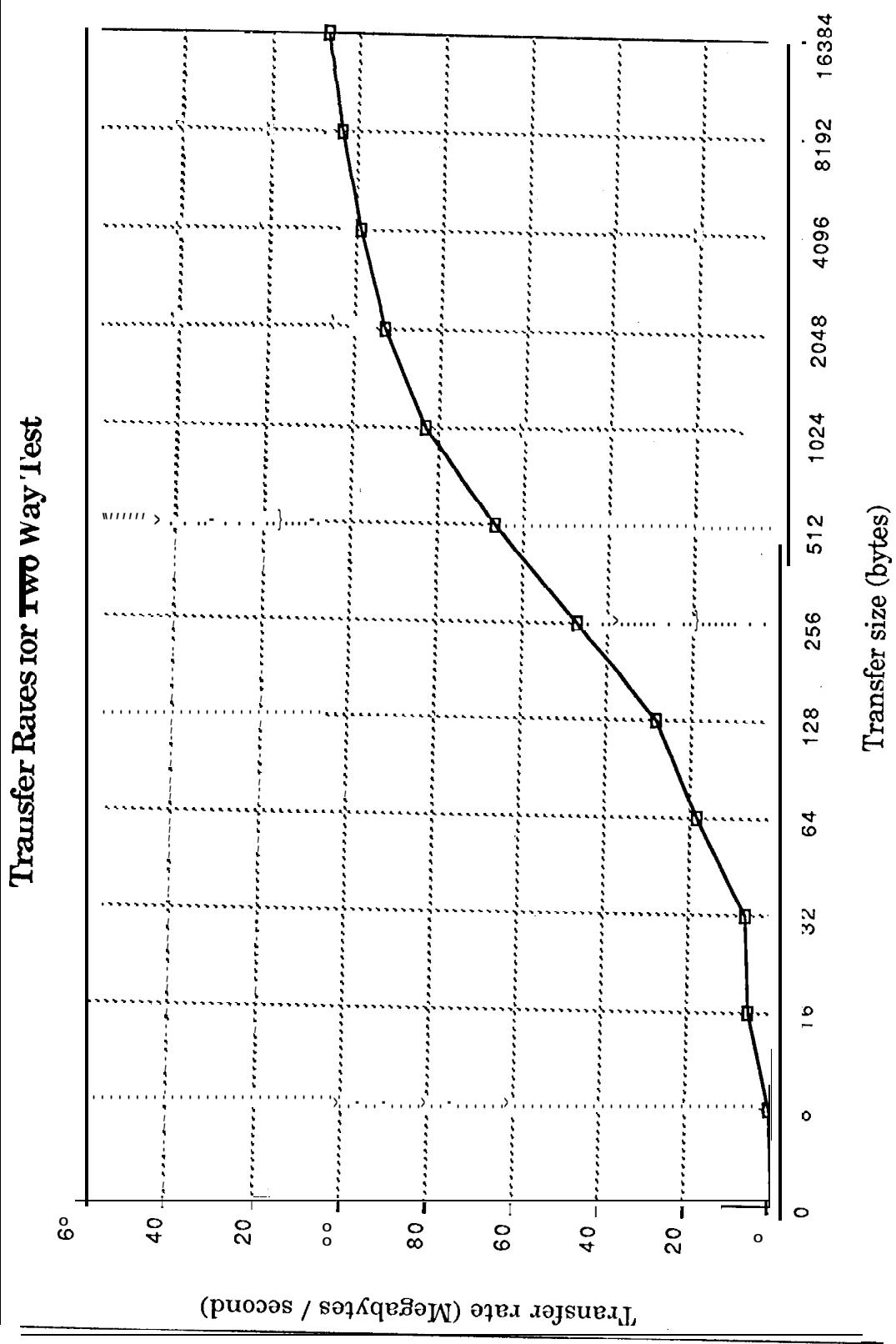
Benchmark Description (continued)

Two Way Test

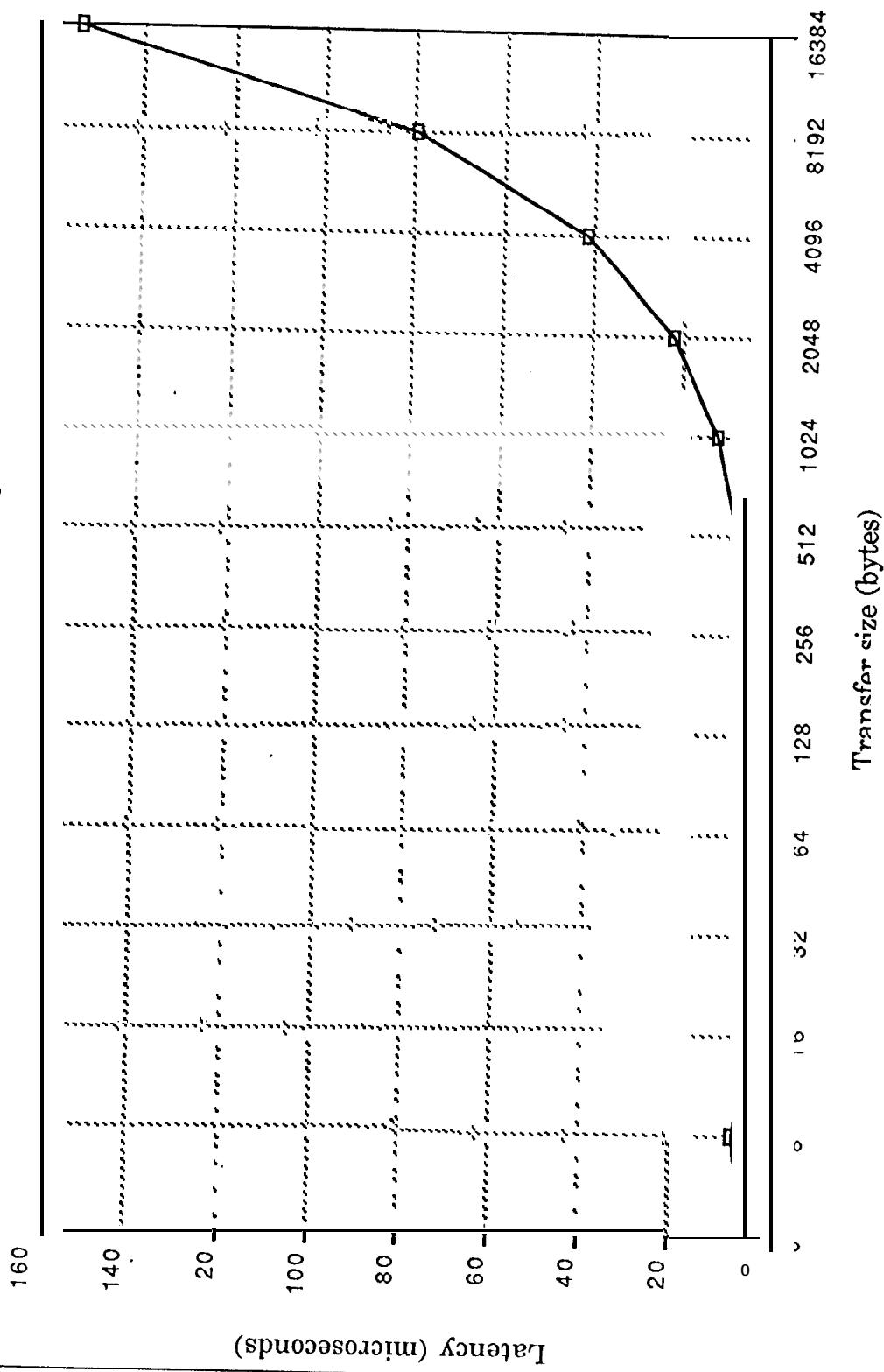
- Two PEs, on adjacent nodes, write to each other's memory
- Minimal contention, between acknowledgment packets and data packets
- Timing figures shown include waiting for acknowledgment packets

Results

- Average transfer rate of 106 megabytes/second for 16K transfers
- Latency times as low as 2.7 microseconds for 16 byte transfer



Latency Times for Two Way Test



Preliminary Measurement of Communication Rates on the Cray T3D Interprocessor Network

Benchmark Description (continued)

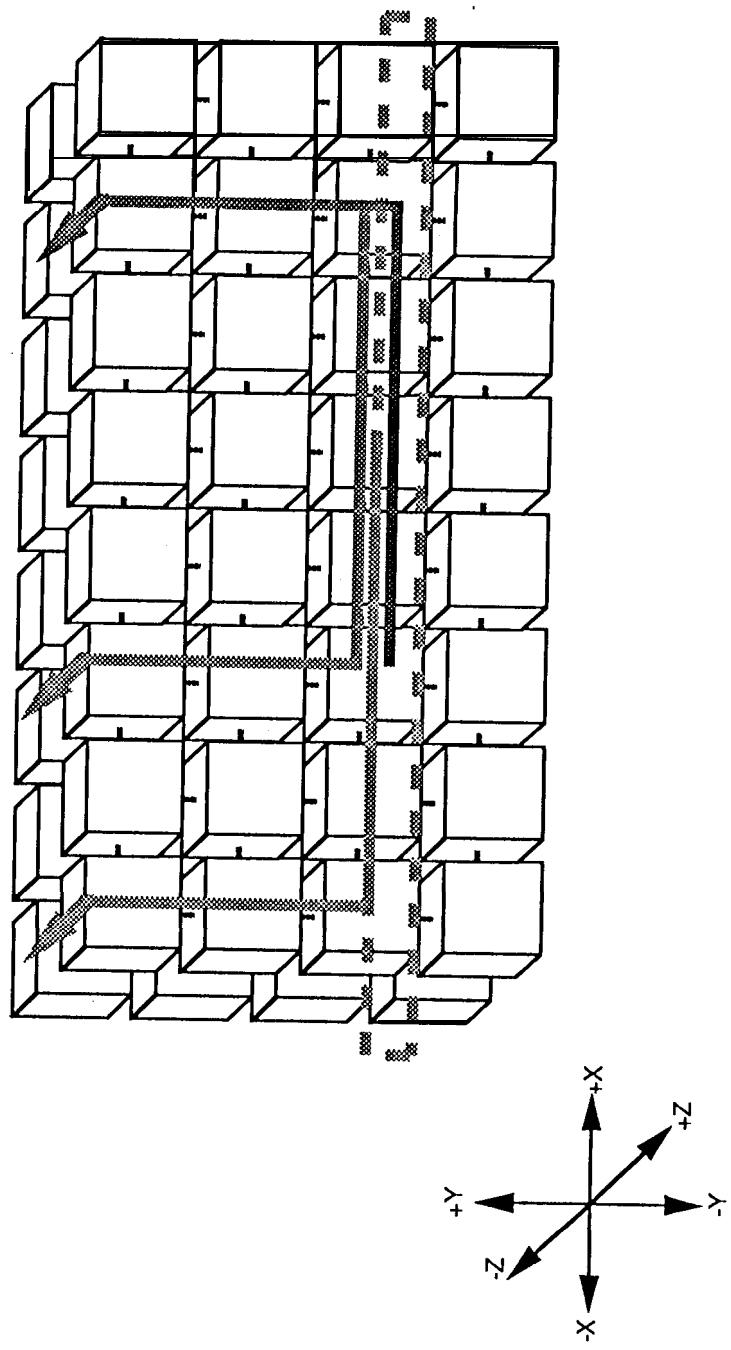
Two Way Tests with Contention

- Each PE is paired with another PE located the maximum distance away
- Each PE in a pair writes to the other's memory
- Maximizes contention
- Timing figures shown include waiting for acknowledgment packets

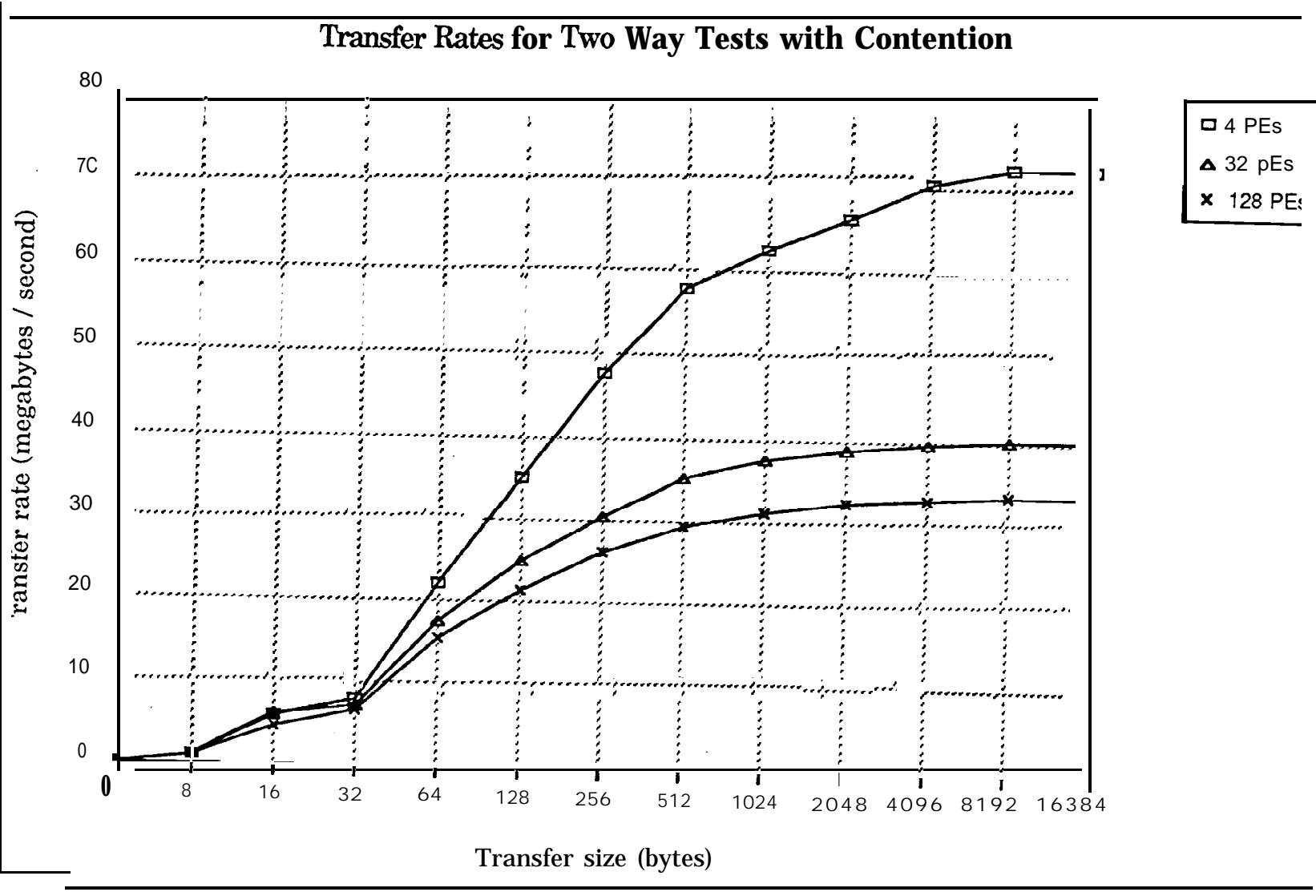
Results

- Average transfer rate of 34 megabytes / second for 16K transfers using 128 PEs in a 8 x 4 x 2 node configuration
- Average latency times of 3.3 microseconds for 16 byte transfer

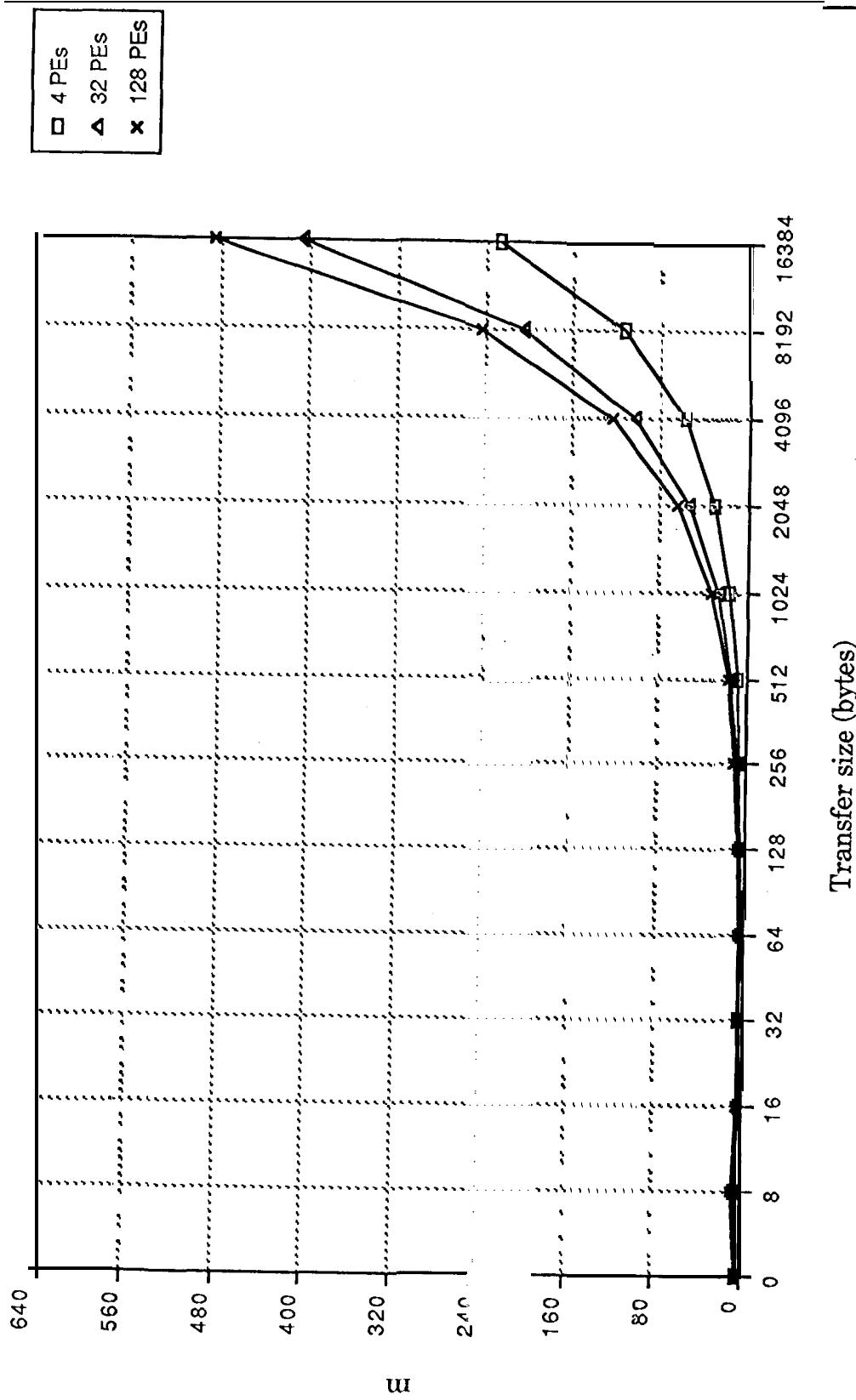
Two Way Test with Contention



Preliminary Measurement of Communication Rates on the Cray T3D Interprocessor Network



Latency Times for Two Way Tests with Contention



Summary

- **Best case transfer rates of 110 megabytes/ second are achievable by an application**
- **Best case latency time is 2.3 microseconds**
- **Early indications are that transfer rates are more than double the rate of previous results, and latency times are an order of magnitude better**
- **Path length has little effect on communication speeds**
- **Heavy contention can reduce transfer rates to 34 megabytes/ second for large transfers**